

Virtual Shelves II: A Unified Catalog for a Heterogeneous Collection

Timothy B. Patrick, Ph.D.^{1,2}, David Moxley, M.A.^{1,2}, Gary K. Allen, D.V.M., Ph.D.^{1,3},
Jean Anderson, M.A.², Shauna Kuster, M.A.²

¹Medical Informatics Research Group, School of Medicine,

²School of Library and Informational Science, ³College of Veterinary Medicine,
University of Missouri-Columbia, Columbia, Missouri

We describe a unified catalog of traditional and digital resources in medical informatics, using standard cataloging principles (AACR2), schemes (LCC), and coding formats (USMARC). The unified catalog integrates the bibliographic records of physical items in the heterogeneous collection with the bibliographic records of network accessible digital items, using prescribed cataloging formats to new effect. The unified catalog is collections-based. We do not use the MARC 856 field to specify the network location of a digital item. The location of a digital item is determined by mapping its call number through a location guide to a network address. This mapping is strictly isomorphic to the way the shelf location of a physical item is determined within the bounds of a controlled collection.

INTRODUCTION

The need to incorporate bibliographic records for digital, computer network accessible items into the traditional library catalog continues to be a focus of concern and discussion in the informatics and library science communities [1-4]. Discussions of these matters have focused on methods for cataloging remote Internet resources. Attempts to use the proposed Machine-Readable Cataloging (MARC) 856 field containing a Uniform Resource Locator (URL) to point to the network location of digital items are well known. A drawback of this use of the MARC 856 field is that URLs for items not controlled by the library must be assumed ultimately unreliable due to the vagaries of remote server and file system management. A URL that is valid one day must not be assumed to be valid the next. In short, where remote Internet resources are not under the direct control of the library, a URL is an unreliable source of location information. Within a controlled collection, however, a URL may be used as a reliable, if somewhat inflexible, source of location information.

Though the MARC 856 field containing a URL may be used successfully to locate digital items in a controlled collection, we reject its use. In this paper we describe a unified catalog for a controlled heterogeneous

collection. The catalog is unified because a uniform set of rules is used to catalog both physical and digital items. We do not treat digital items, per cataloging requirements, as fundamentally different in kind from physical items and so we do not use the MARC 856 field as a source of location information for digital items in the collection. The location of digital items in the collection is determined in a manner strictly isomorphic to the way physical items are located. No extra location field is needed for digital items within the collection. The approach described here is an elaboration of the method previously described in [5].

METHOD

Figure 1 shows the MARC and public display records for two items in our heterogeneous collection. Both items are copies of *Information Retrieval* by C. J. Van Rijsbergen. The MARC and public display record on the left of Figure 1 are for a physical copy of *Information Retrieval*. According to its *Call number* (LCC) field 090, the physical copy is located in the University of Missouri-Columbia (MU) campus branch library ELLIS. The MARC and public display record on the right of Figure 1 are for a digital copy of *Information Retrieval*. According to its *Call number* field 090, the digital copy is located in the HTTP region of the MU campus library intranet.

The digital copy of *Information Retrieval* is not treated, per cataloging requirements, as fundamentally different in kind from the physical copy. The apparently striking difference in the *Call number* field 090, viz., HTTP versus ELLIS, is unimportant. From the point of view of cataloging, HTTP may be considered a branch library similar to ELLIS. Thus the catalog for this heterogeneous collection is *unified* in the sense that one uniform set of rules is applied to both physical and digital items.

It remains for us to show, however, that this uniformity is justified. In this paper we address the especially important issue of *location*—whether the information contained in the catalog records may be used reliably to locate the corresponding copies of *Information*

MARC Record						MARC Record					
Type: a	Elvl:	Src:	Audn:	Ctrl:	Lang: eng	Type: a	Elvl:	Src:	Audn:	Ctrl:	Lang: eng
Blvl: m	Form:	Conf: 0	Biog:	Mrec:	Ctry: enk	Blvl: m	Form:	Conf: 0	Biog:	Mrec:	Ctry: enk
	Cont: b	Gpub:	Fict: 0	Indx: 1			Cont: b	Gpub:	Fict: 0	Indx: 1	
Desc: a	Ills: a	Fest: 0	DtSt: s	Dates: 1975,		Desc: a	Ills: a	Fest: 0	DtSt: r	Dates: 1995, 1975	
010	75-33024					024 8	MU000002				
020	0408707178					090 0	HTTP #a Z699 #a .V35				
024 8	MU000001					100 1	Van Rijsbergen, C. J., #d 1943-				
090 0	ELLIS #a Z699 #a .V35					245 10	Information retrieval / #c C. J. van Rijsbergen.				
100 1	Van Rijsbergen, C. J., #d 1943-						#h [computer file]				
245 10	Information retrieval / #c C. J. van Rijsbergen.					260	London ; #a Boston : #b Butterworths, #c 1995, 1975				
260	London ; #a Boston : #b Butterworths, #c 1975.					300	[ca. 1018MB] : #b ill..				
300	152 p. : #b ill. ; #c 23 cm.					500	Document title				
500	Includes index.					500	Approximately 152 pages when printed.				
650 0	Information storage and retrieval systems.					500	HTML document				
650 0	Automatic indexing.					650 0	Information storage and retrieval systems.				
650 0	Classification.					650 0	Automatic indexing.				
						650 0	Classification.				
Public Display Record						Public Display Record					
AUTHOR	Van Rijsbergen, C. J. 1943-					AUTHOR	Van Rijsbergen, C. J. 1943-				
TITLE	Information Retrieval/C. J. van Rijsbergen					TITLE	Information Retrieval [computer file]				
PUBLISHER	London; Boston ; Butterworths ; 1975						/C. J. van Rijsbergen				
DESCRIPTION	152 p. : ill. ; 23 cm.					PUBLISHER	London; Boston ; Butterworths ; 1995, 1975				
NOTES	Includes index					DESCRIPTION	[ca. 1.018MB] : ill.				
SUBJECTS	Information storage and retrieval systems					NOTES	Approximately 152 pages when printed.				
	Automatic indexing						HTML document.				
	Classification					SUBJECTS	Information storage and retrieval systems				
							Automatic indexing				
							Classification				
<u>LOCATION</u>	<u>CALL#</u>					<u>LOCATION</u>	<u>CALL#</u>				
ELLIS	Z699.V35	(1 st floor west)				HTTP	Z699.V35				

Figure 1 - MARC and public display records for physical and digital copies of *Information Retrieval*

Retrieval. The information contained in the MARC record for the physical copy of *Information Retrieval* is clearly sufficient, given traditional library practices, to allow us to locate it. If the information contained in the MARC record for the digital copy of *Information Retrieval* is not so sufficient, or does not lend itself to *location* solutions that are scalable across the MU library enterprise, then we have evidence that our approach to cataloging the digital copy is defective. In that case it will be reasonable to conclude that we require an approach to cataloging that treats digital items as fundamentally different in kind from physical items.

Fortunately, however, the information contained in the MARC record for the digital copy of *Information Retrieval* is sufficient to allow us reliably to locate the digital copy, and to do so in a way that is scalable across the MU library enterprise. To be sure, the processes used to locate and ultimately access the physical and digital copies must differ in detail.

However, the steps in those respective processes are strictly isomorphic.

In order to locate the physical copy of *Information Retrieval*, we use the subject classifier portion of the call number, Z699, as an indirect pointer to a physical shelf in a physical room of the MU branch library ELLIS. Following a familiar pattern, Z699 is matched against a shelf location guide that indicates the room. Furthermore, Z699 is matched against indicators on the ends of shelves in that room. The Cutter portion of the call number, .V35, is used to fix relative position on the appropriate shelf.

Similarly, in order to locate the digital copy of *Information Retrieval*, we use the subject classifier portion of the call number, Z699, as an indirect pointer to a virtual shelf [5] in the HTTP portion of the MU library intranet. A virtual shelf is a server on a particular host computer that stores digital items for a given range of subject classes. Z699 is matched against

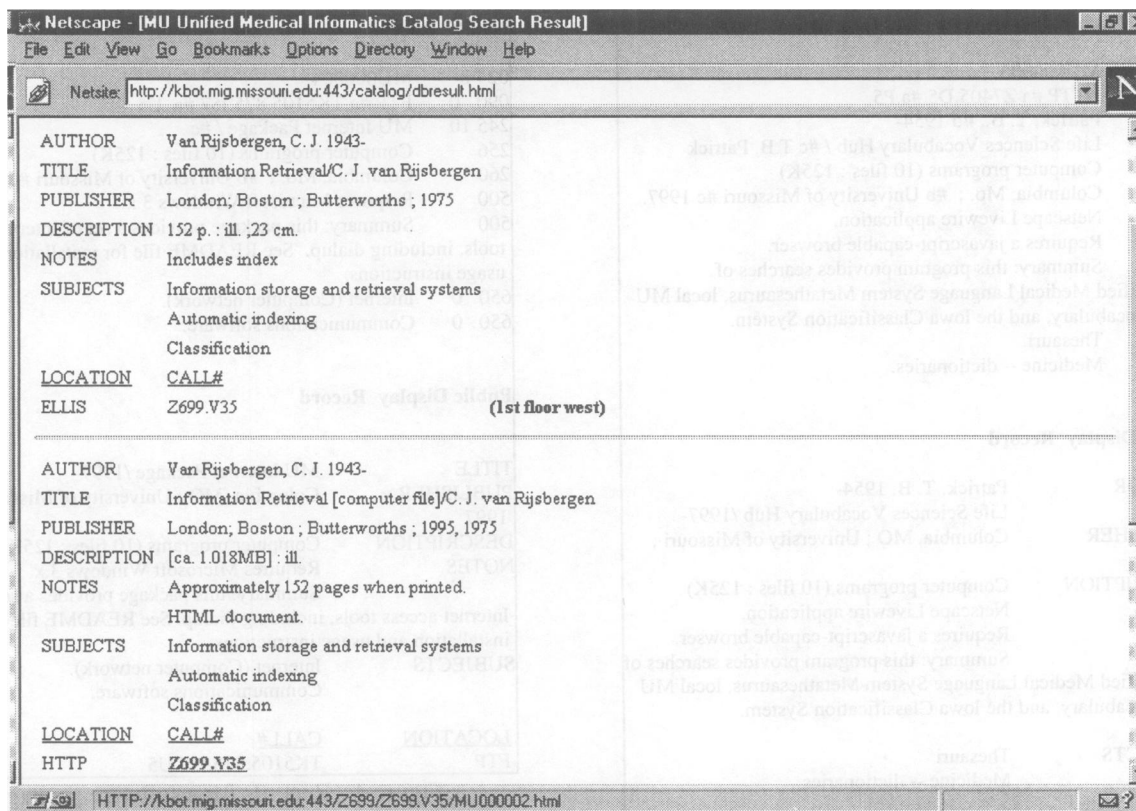


Figure 2 - Catalog search result

a location guide that indicates the virtual shelf, i.e., the server. Furthermore, Z699 is used to indicate a particular subdirectory in the file system of the server. The Cutter portion of the call number, .V35, is used to fix relative position in the subdirectory Z699.

Figure 2 depicts the results of a sample search of our catalog. In the case of the physical copy of *Information Retrieval*, the catalog search system automatically used an on-line shelf location guide to determine and report the location of the physical room in ELLIS where the physical copy can be found, viz., 1st floor west.

In the case of the digital copy of *Information Retrieval*, our catalog search system uses the information in the MARC record and the shelf location guide to automatically construct a URL for the digital copy. The "branch library name", HTTP, is used as the protocol type. The location guide indicates a server, e.g. kbot.mig.missouri.edu:443. Z699 identifies a particular subdirectory on the server, and .V35 indicates a particular subdirectory of Z699. Since it is unique for digital items within the bounds of our catalog, we could use the call number as the basis for the filename of the digital item. However, insuring

uniqueness requires that the full call number may be rather lengthy, and we choose not to use the call number as the basis for the filename. For purposes of URL readability we instead use the value of the *Other Standard Identifier* field 024 as the basis for the filename. All items in our catalog are assigned local MU identifiers with the 024 field. The identifier for the physical copy of *Information Retrieval* is MU000001; the identifier for the digital copy is MU000002. MU000002 is used to construct the filename of the digital copy. The resulting URL is

HTTP://kbot.mig.missouri.edu:443
/Z699/Z699.V35/MU000002.html

The constructed URL is visible in the status window of the browser.

The digital item cataloged in Figure 1 is a static document. However, given the flexibility of the HTTP protocol, we can easily use the same approach to include executable programs in our catalog. Figure 3 shows the MARC and public display records for an on-line vocabulary search system. The URL constructed from the MARC record is

MARC Record	
024 8	MU000003
090 0	HTTP #a Z7405.D5 #a.P5
100 1	Patrick, T. B., #d 1954-
245 10	Life Sciences Vocabulary Hub / #c T.B. Patrick
256	Computer programs (10 files : 125K)
260	Columbia, Mo. ; #b University of Missouri #c 1997.
500	Netscape Livewire application.
500	Requires a javascript-capable browser.
500	Summary: this program provides searches of the Unified Medical Language System Metathesaurus, local MU HSC vocabulary, and the Iowa Classification System.
650 0	Thesauri.
650 0	Medicine -- dictionaries.
Public Display Record	
AUTHOR	Patrick, T. B. 1954-
TITLE	Life Sciences Vocabulary Hub /1997-
PUBLISHER	Columbia, MO ; University of Missouri ; 1997
DESCRIPTION	Computer programs (10 files : 125K)
NOTES	Netscape Livewire application. Requires a javascript-capable browser. Summary: this program provides searches of the Unified Medical Language System Metathesaurus, local MU HSC vocabulary, and the Iowa Classification System.
SUBJECTS	Thesauri Medicine -- dictionaries
<u>LOCATION</u>	<u>CALL#</u>
HTTP	Z7405 .D5 .P5

Figure 3 - Records for an executable program

HTTP://kbot.mig.missouri.edu:443
/Z7405.D5/Z7405.D5.P5/MU000003.html

Our choice of cyberspace "branch library" is not restricted to HTTP. Any of the usual Web protocols are available. For example, we may choose to use the HTTP "branch library" to hold executable programs, and we may choose to use the FTP "branch library" to hold software packages available for installation. Figure 4 shows the MARC and public display records for the MU Internet package. The URL constructed from the MARC record is

FTP://condor.cs.missouri.edu
/TK5105.875.I57/
TK5105.875.I57.U5/MU000004.exe

CONCLUSION

We have shown that, though the MARC 856 field may be used successfully to locate digital items in a controlled collection, it is not needed. There are in addition other good reasons for rejecting the MARC

MARC Record	
024 8	MU000004
090 0	FTP #a TK5105.875.I57 #a .U5
245 10	MU Internet Package / #c
256	Computer programs (10 files : 125K)
260	Columbia, Mo. ; #b University of Missouri #c 1997.
500	Requires Microsoft Windows 3.x.
500	Summary: this package provides a set of Internet access tools, including dialup. See README file for installation and usage instructions.
650 0	Internet (Computer network).
650 0	Communications software.
Public Display Record	
TITLE	MU Internet Package /1997-
PUBLISHER	Columbia, MO ; University of Missouri ; 1997
DESCRIPTION	Computer programs (10 files : 125K)
NOTES	Requires Microsoft Windows 3.x. Summary: this package provides a set of Internet access tools, including dialup. See README file for installation and usage instructions.
SUBJECTS	Internet (Computer network). Communications software.
<u>LOCATION</u>	<u>CALL#</u>
FTP	TK5105.875.I57.U5

Figure 4 - Records for a software package

856 field.

One reason for rejecting the MARC 856 field is that a URL is an inflexible source of location information. According to our method, the location information for an entire range of digital items may be updated by making a single change in a field in the library location directory. However, if digital items are cataloged using the MARC 856 field containing a fixed URL, updating the location information for digital items will require us to update individual catalog records.

A second more important reason for rejecting the MARC 856 field concerns the very essence of cataloging. The purpose of cataloging is to catalog the *item in hand*. In a digital context, having the item in hand means having the item under the library's control. The essential point is that control of the item and its bibliographic surrogate must reside in the same hands.

Consider that our model assumes local control of every item represented in the unified catalog, as well as obvious control over the surrogates. The MARC 856 field seems designed to deal with items that are not under local control, although the bibliographic surrogate is. We argue that this use of the MARC 856 field is an unnecessary and inadvisable artifice. The appropriate way to locate a remote digital item is to use

the remote catalog for the controlled collection containing that item. In order to provide access to items outside of the collection under local control, a network of catalogs could be established across multiple institutions. Each institution in the network would exercise exclusive control of the items and surrogates in their individual collections. End users could search multiple catalogs in the network, and be assured of locating a selected item within its respective collection. Situations in which control of items in a collection is divorced from control of their bibliographic surrogates are excluded from this model. For us, control of the bibliographic surrogate necessarily implies control of the item.

Acknowledgments

This work was supported in part by grants LM07089 and LM05415 from the National Library of Medicine.

References

1. Brugger JM. Cataloging for digital libraries. *Cataloging & Classification Quarterly* 1996; 22(3-4):59-73.
2. Dillon M, Jul E. Cataloging Internet resources: the convergence of libraries and Internet resources. *Cataloging & Classification Quarterly* 1996; 22(3-4):197-238.
3. Morgan EL. Possible solutions for incorporating digital information mediums into traditional library cataloging services. *Cataloging & Classification Quarterly* 1996; 22(3-4):143-170.
4. Mandel CA, Wolven R. Intellectual access to digital documents: joining proven principles with new technologies. *Cataloging & Classification Quarterly* 1996; 22(3-4):25-42.
5. Patrick TB, Springer GK, Mitchell JA, Sievert ME. Virtual shelves in a digital library: a framework for access to networked information sources. *JAMIA* 1995; 2(6): 383-390.